Genetic Risk Scores for Complex Disease Traits in Youth

Xie, Tian; Wang, Bin; Nolte, Ilja M; van der Most, Peter J; Oldehinkel, Albertine J; Hartman, Catharina A; Snieder, Harold

Published in:
Circulation. Genomic and precision medicine

DOI:
10.1161/CIRCGEN.119.002775

Publication date:
2020

Genetic Risk Scores for Complex Disease Traits in Youth

Running title: *Xie et al.; GRSs for complex disease traits in youth*

Tian Xie, MSc[1]; Bin Wang, PhD[1]; Ilja M. Nolte, PhD[1]; Peter J. van der Most, PhD[1]; Albertine J. Oldehinkel, PhD[2]; Catharina A. Hartman, PhD[2]; Harold Snieder, PhD[1]

[1]Department of Epidemiology, [2]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Correspondence:
Harold. Snieder, PhD
Department of Epidemiology
University Medical Center Groningen
Hanzeplein 1
PO Box 30001
9700 RB Groningen,
The Netherlands
Tel +31 50 361 0887
Fax +31 50 36 14493
E-mail h.snieder@umcg.nl

Journal Subject Terms: Genetic, Association Studies; Primary Prevention; Cardiovascular Disease

Abstract:

Background - For most disease related traits the magnitude of the contribution of genetic factors in adolescents remains unclear.

Methods - Twenty continuous traits related to anthropometry, cardiovascular and renal function, metabolism and inflammation were selected from the ongoing prospective TRAILS (TRacking Adolescents' Individual Lives Survey) cohort in the Netherlands with measurements of up to five waves from age 11 to 22 years (n=1,354, 47.6% males) and all traits available at the third wave (mean age [SD] = 16.22 [0.66]). For each trait, unweighted and weighted genetic risk scores (GRSs) were generated based on significantly associated SNPs identified from literature. The variance explained by the GRSs in adolescents were estimated by linear regression after adjustment for covariates.

Results - Except for alanine transaminase (ALT), all GRSs were significantly associated with their traits. The trait variance explained by the GRSs was highest for lipoprotein[a] (39.59%) and varied between 0.09% (ALT) and 18.49% (low-density lipoprotein (LDL)) for the other traits. For most traits the variances explained in adolescents were comparable with or slightly smaller than those in adults. Significant increases of trait levels (except ALT) and increased risks for overweight/obesity (OR 6.41, 95% CI 2.95-15.56) and hypertension (OR 2.86, 95% CI 1.39-6.17) were found in individuals in the top GRS decile compared with those at the bottom decile.

Conclusions - Variances explained by adult-based GRSs for disease related traits in adolescents, although still relatively modest, were comparable with or slightly smaller than in adults offering promise for improved risk prediction at early ages.

Key words genetics, human; young; risk score; genetic risk score; complex disease traits; variance explained

Nonstandard Abbreviations and Acronyms

ALT	alanine transaminase
BMI	body mass index
CRP	C-reactive protein
DBP	diastolic blood pressure
eGFR	estimated glomerular filtration rate
FG	fasting glucose
FGadjBMI	fasting glucose (BMI adjusted)
FI	fasting insulin
FIadjBMI	fasting insulin (BMI adjusted)
GRS	genetic risk score
HbA1c	glycated hemoglobin
HDL	high-density lipoprotein
HR	heart rate
IgE	Immunoglobulin E
IOTF	The International Obesity Task Force
LD	linkage disequilibrium
LDL	low-density lipoprotein
Lp(a)	lipoprotein(a)
MAF	minor allele frequency
meta-GWASs	meta-analyses of genome-wide association studies
PRS	polygenic risk score
SBP	systolic blood pressure
SNP	single nucleotide polymorphism
TC	total cholesterol
TG	triglycerides
TRAILS	TRacking Adolescents' Individual Lives Survey
TRAILSCC	TRAILS clinical cohort
uGRS	unweighted genetic risk score
wGRS	weighted genetic risk score
WHRadjBMI	waist-to-hip ratio (BMI adjusted)

Introduction

Recently, an increasing number of genetic variants - mostly single nucleotide polymorphisms (SNPs) - have been identified to be associated with human traits through meta-analyses of genome-wide association studies (meta-GWASs).[1,2] To evaluate the overall contribution of these identified genetic variants, genetic risk scores (GRSs) were constructed for many disease related traits and were found to often explain a significant portion of the trait variation.[3] As more SNPs continue to be discovered, such GRSs provide possibilities to predict complex disease risk at the individual level and have potential application in disease prevention.

For example, many studies on blood pressure and body mass index (BMI) have shown that increased levels in youth track into adulthood and are associated with immediate and long-term health risks.[4-7] Applying GRSs at an early age to identify individuals at high genetic risk for hypertension and obesity might therefore aid in early prevention. As most SNPs were identified from meta-GWAS in adults, the question whether these adult based GRSs can be applied in youth needs to be answered. A longitudinal twin study on blood pressure showed that novel genetic effects emerged between ages 14 and 18 years and explained a significant part of the variation in blood pressure.[8] Another study found that five loci had different effects on BMI during adolescence and young adulthood (16-25 years) compared with middle-age adults.[9] These results support age-dependent genetic effects and suggest that GRSs derived from adults may not have the same effect in youth. Thus, there is the need to investigate to what extent adult based GRSs can predict disease related traits in youth.

However, so far only a few traits and diseases were explored, and the contributions of GRSs in adolescents for other traits remain unclear.[10-12] Furthermore, as the list of identified

genetic markers has recently expanded dramatically, the effect of updated GRSs using the latest GWAS findings requires evaluation in adolescents.

Therefore, the aim of the current study was to assess and evaluate the variance explained by adult based GRSs on a wide variety of disease related traits in adolescents from the Netherlands. We used 20 continuous traits from the TRacking Adolescents' Individual Lives Survey (TRAILS) cohort, related to anthropometry, cardiovascular and renal function, metabolism and inflammation. For each trait we generated GRSs based on significantly associated SNPs identified from literature. Then we assessed how much of the phenotypic variance could be explained by these GRSs in adolescents and compared it with the phenotypic variance explained in adult populations. We also compared the trait levels and risks of hypertension and obesity between individuals in each of the upper nine deciles with those in the bottom decile of the GRSs distribution. Furthermore, we replicated findings of some major traits such as BMI and blood pressure in the TRAILS clinical cohort (TRAILSCC).

Methods

The research was conducted in TRAILS, an ongoing prospective population-based cohort which assesses physical and psychosocial health from preadolescence to adulthood in the Netherlands.[13,14] Because of the personal nature of the data, the dataset is not online available. Requests to access the data may be submitted by means of a publication plan form for external users, which is available at https://www.trails.nl/en/hoofdmenu/data/data-use.

The traits of interest and the literature from which SNPs were identified are presented in Table 1. All SNPs and their effect sizes for constructing GRS can be found in Supplementary

Table 1-21. Full descriptions of trait and SNP selection, participants and traits measurements, genotyping and imputation, and statistical analyses are available in the Supplementary Methods.

All procedures were approved by the Dutch Central Committee on Research Involving Human Subjects. Written informed consent, including specific consent to undertake genetic analyses, was obtained from participants and their parents or custodians.

Results

Participants and traits description

Table 2 shows the descriptive statistics of age and the quantitative traits we selected in the TRAILS cohort at the third wave (for all waves see Supplementary Table 22). A total of 1,354 participants whose GWAS data were available were included in the analyses, 644 (47.6%) of whom being males. The mean ages (in years) of the five waves (T1 – T5) were 11.1, 13.5, 16.2, 19.2 and 22.4, respectively. In total 20 traits, related to anthropometry, cardiovascular and renal function, metabolism and inflammation were selected.

SNPs selection

Figure 1 shows the process and results of SNP selection for the 20 traits of interest. We selected 17 articles as sources of SNPs for the 20 traits, of which 13 used GWAS data, 2 used exome-centric chips, and 2 used a combination of GWAS and gene-centric data. From these papers, we identified 8,183 SNP-phenotype associations. For 35 associations SNPs were missing in the TRAILS genotyped or imputed data, but we could successfully find proxies for 10 of them. Eighty one associations were removed because SNPs were in LD with another selected SNP. Finally, 8,077 SNP-phenotype combinations were included for constructing GRSs of the 20 traits (Figure 1; Supplementary Table 1-21).

Genetic risk score analysis

Table 3 shows the results of weighted genetic risk score analysis at the third wave (for results of uGRS see Supplementary Table 23, for all waves see Supplementary Table 24). The number of SNPs included in the GRSs ranged from 4 (for ALT) to 3,290 (for height). Except for the GRSs for ALT, all GRSs were significantly associated with their traits and explained a significant part of the phenotypic variance. The variance explained by the GRSs for the traits varied greatly: the weighted GRS incorporating 49 SNPs for lipoprotein(a) (Lp(a)) explained 39.59% of its variance, while the wGRS for ALT only explained 0.10% and was not significant. Apart from Lp(a), GRSs for height, high-density lipoprotein (HDL), LDL and total cholesterol (TC) had relatively large contributions to their traits (above 10%). For blood pressure, the variance explained by the wGRS is 2.15% for SBP and 4.48% for DBP. For anthropometric traits that had repeated measurements (Figure 2, Supplementary Table 24), we found increases in variance explained by the GRSs for height with older age (e.g. from 9.34% at 14 years of age to 12.03% at the age of 16 for the uGRS), but the differences were not significant. The variances explained remained similar for BMI from 11 to 22 years (between 5.79% and 6.55% for wGRS) and for WHRadjBMI from 16 to 22 years (between 1.38% and 1.95% for wGRS).

The variance explained by GRSs increased when using the wGRSs compared to the uGRSs for most traits (Supplementary Table 23, Supplementary Figure 1), with the biggest increase for LDL. The uGRS for LDL explained 8.88% of the variance compared with 18.49% by the wGRS, an increase of almost 10%. For the traits that needed correction for 'the TRAILS' result, wGRSs using corrected effect sizes explained slightly less variance (range: 0.01% - 0.43%) than wGRSs using uncorrected effect sizes from the literature as expected (Supplementary Table 25).

For 17 traits, variances explained in adolescents were compared with variance explained in adults that were extracted from literature (Figure 3, Table 3 and Supplementary Table 26). Generally, variances explained in adolescents were similar or slightly less than those in adults, with the biggest difference for C-reactive protein (CRP). The variance explained for CRP was 3.69% in adolescents compared with 11% in adults.

For all traits except ALT, significant increases of trait levels were found in individuals at the top wGRS decile compared with those at the bottom decile (Table 3). For instance, individuals at the top decile of the wGRS for SBP had on average a 6.30 mmHg higher SBP (95% CI 3.54 to 9.07 mmHg) than those at the bottom decile. For most traits levels of trait increased along with increases in the GRS decile (Supplementary Figure 2). Furthermore, over six-fold higher risk of overweight/obesity (OR 6.41, 95% CI 2.95-15.56) and around three-fold higher risk of hypertension (OR 2.86, 95% CI 1.39-6.17) were observed between top and bottom deciles of the GRS (Figure 4).

Replication in TRAILSCC

From TRAILSCC, 341 participants (69.2% males) with available DNA were included in replication analyses for height, BMI, WHRadjBMI, HR, SBP, DBP, HbA1C(%). Generally, GRSs also explained significant proportions of phenotypic variance in TRAILSCC (Supplementary Table 27).

Additional analyses for BMI and blood pressure

Fifteen SNPs were identified in meta-GWAS of childhood BMI and two SNPs for SBP in children or adolescents.[32,33] For childhood BMI, six SNPs showed significant associations with BMI at 11 years in TRAILS (P<0.05) and 14 SNPs had directionally consistent effects with those reported by meta-GWASs (Supplementary Table 28). For childhood SBP, the two SNPs

were not significantly associated with SBP in TRAILS, but they had the same direction of effect as those in the meta-GWAS (Supplementary Table 28). Comparing with the uGRSs only including SNPs identified in adults, the uGRSs combining SNPs identified in adults and in children/adolescents explained slightly more variance of SBP and BMI at 11 years and 14 years (e.g. SBP, R^2=1.69% compared to 1.67%) (Supplementary Table 29).

Discussion

In this study, we investigated in 1,354 Dutch adolescents how much of the variance of 20 complex disease traits could be explained by adult based GRSs. Our results showed that almost all adult based GRSs were significantly associated with their respective traits in adolescents. The trait variance explained by the GRSs varied from 0.09% to 39.59%, with weighted GRSs generally explaining a larger proportion of variance than the unweighted GRSs. For most traits the variance explained in adolescents was comparable with or slightly less than in adults. Significant increases of trait levels (except ALT) and increased risks for overweight/obesity and hypertension were found in individuals in the top wGRS decile compared with those at the bottom decile.

Among metabolism traits, the variance explained by GRSs varied greatly, which may be caused by the differences in trait heritabilities and genetic architecture. For example, the wGRS for ALT explained 0.01% of variance which was not significant, while the wGRS for Lp(a) explained nearly 40%. The small variance explained for ALT is likely due to moderate heritability (22%-40% estimated from twin-family studies) and the GRS including only 4 SNPs as many ALT-associated SNPs may not yet have been identified due to insufficient power of the discovery GWAS.[34,35] For some other liver enzymes, such as alkaline phosphatase (ALP) and γ-

glutamyl transferase (GGT), more SNPs were identified in the original GWAS, but these traits were not measured in TRAILS.[22] On the other hand Lp(a) is highly heritable (~90%), with 48 identified SNPs located in the *LPA* gene region and only one SNP in another gene (*APOE*), indicating that this trait is not (very) polygenic in its architecture.[29,36,37] For lipid traits (HDL, LDL, TC and Triglycerides (TG)), we selected SNPs from both genome-wide and exome-centric association studies including some rare variants (minor allele frequency (MAF) <1%). We found that the GRSs excluding rare variants explained slightly less variance of the lipid traits than the GRSs including rare variants, which indicates that even these rare variants with low imputation quality in TRAILS contribute to lipid trait variance (Supplementary Table 30-31). For repeatedly measured traits, no significant change was found between different waves. This is probably due to insufficient power of our sample or relatively stable influences of genetic factors during this age period (11 to 22 years).

The associations between adult based GRSs and their respective traits in adolescents suggest that many of SNPs identified in adults also have effects in adolescents. Similar findings were reported before for GRSs based on blood pressure and BMI loci.[10,12] Another study found a genetic correlation of 0.73 between childhood and adult BMI as calculated by LD score regression, indicating large but not perfect genetic overlap between childhood and adult BMI.[32] These results indicate the potential of applying adult-based GRSs to disease related traits for prediction at an early age. Besides, additional analyses for BMI and blood pressure suggested that combining SNPs identified in adults and in children/adolescents can increase the predictive ability of GRS. If more SNPs will be identified in future GWASs of children or adolescents, GRSs of these traits will likely explain more variance of their traits. In addition, Evangelou et al. discovered that the wGRS for blood pressure was associated with increased risk of

cardiovascular events during adulthood.[19] Our study showed that the adult-based GRSs for blood pressure and BMI could also predict hypertension and overweight/obesity, respectively, during adolescence. Therefore, GRSs for blood pressure and BMI may have the potential to guide preventative measures for hypertension and obesity in youth. For example, lifestyle interventions such as diet and physical activity could be targeted in individuals who are identified at high genetic risk already in early life.

Furthermore, we found that the effects of adult based GRSs are similar or slightly smaller in adolescents compared with adults. The similarities of effects between adolescents and adults suggest that for some traits the influence of genetic factors may remain relatively stable from adolescence to adulthood. One reason for the small differences of effects between adolescents and adults may be age-dependent genetic effects (different genes may play a role or the magnitude of effects of the same genes on the phenotypes may change over time). For instance, variance explained for SBP was less in adolescents than in adults (2.15% vs 5.70%).[19] Other studies on blood pressure confirmed that not all individual SNPs identified in adults were significantly associated with BP in adolescents and that adult-based GRSs explained less variance in adolescents.[12,33] Another reason may be lack of stability of phenotypes for some traits. Levels of some traits during adolescence may be quite different from levels in adulthood as adolescence is a period of rapid anthropometric change. The growth rate varies between individuals as indicated by the less than perfect tracking (e.g., the correlation between height at 11 and 22 years old was only 0.414) indicating that during adolescence individuals may be at different stages of development (Supplementary Table 32). Finally, we observed a much smaller explained variance for CRP in adolescents compared to adults. However, the original meta-GWAS in adults applied the formula $2*MAF(1-MAF)b^2/var$ to estimate variance explained rather

than performing an out-of-sample prediction in an independent validation cohort, which may have caused an overestimate of their explained CRP variance in adults.[30] For some traits our results were not completely comparable with those from literature, as they used different methods or included not exactly the same SNPs as we did to estimate variance explained in adults (Supplementary Table 26). Nonetheless, the comparisons indicate that for most traits genetic markers identified in GWASs of adults may explain similar or slightly less variance in adolescents than in adults.

One statistical issue is that in spite of testing GRSs for twenty different traits we have chosen not to apply a multiple testing correction in our study, because our aim was to replicate significant results from previous studies. We simply provided exact *p* values of GRSs for all traits in Table 3. However, it is important to point out that the GRSs for most traits would remain significant even if we used a corrected significance threshold of 0.0025 for 20 independent tests. Therefore, the interpretation of the results would not change if we would have adjusted for multiple testing.

A limitation of our study is that we included only known genome-wide significant SNPs in our GRSs, instead of using approaches which include all available SNPs like a polygenic risk score (PRS) or LDpred. PRS is similar to GRS but includes larger number of independent SNPs by using more lenient significance thresholds.[38] LDpred includes all SNPs below a certain significance threshold and accounts for linkage disequilibrium (LD) among SNPs to reduce loss of information.[39] Conducting PRS or LDpred requires GWAS summary statistics that were not available for all twenty traits we investigated, so we chose to use the GRS approach, which only needs a list of significant SNPs as published in the literature. As such, our results are conservative; polygenic scores generated by PRS or LDpred are likely to explain more variance.

In addition, we selected only SNPs identified from European ancestry and applied GRSs in adolescents of the same ancestry. Our results may not be applicable to adolescents from other ethnicities.

Despite these limitations, our research contributes to the understanding of genetic influence on twenty traits during a specific life period. To our knowledge, we are the first to evaluate the variance explained by adult-based GRSs for a wide range of disease-related traits in one homogeneous adolescent cohort. Even with a relatively small sample size, we detected associations between adult based GRSs and 19 traits in adolescents. In addition, as we had repeated measurements for some traits such as height, BMI and WHR, we could observe the contributions of known SNPs to these traits at different ages during the critical time period from childhood to early adulthood. Further, we calculated GRSs using the latest GWAS findings, so we could evaluate the value of applying updated adult based GRSs in adolescence. With the help of larger GWAS studies, more GWASs in children or adolescents and improved approaches for calculating genetic predictors, genetic risk prediction is likely to further gain accuracy. As genetic predictors can be calculated for many diseases simultaneously from birth onwards, genetic risk prediction provides opportunities to identify high-risk strata for many diseases at an early age, which is especially important for diseases with known effective interventions.

In conclusion, we demonstrated that almost all adult based GRSs for 20 continuous disease trait were significantly associated with their respective traits in adolescents. Overall, these adult based GRSs explained a small to moderate part of phenotypic variance in adolescents and their effects appeared comparable with or slightly smaller than in adults. Larger GWAS studies and improved approaches to calculating genetic predictors in combination with efforts to

integrate genetic, environmental, clinical and molecular risk factors may offer promise for improvement of disease prevention.

Acknowledgments: This research is part of the Tracking Adolescents' Individual Lives Survey (TRAILS). We are grateful to everyone who participated in this research or worked on this project to make it possible.

Sources of Funding: Tian Xie was financially supported by a grant from the China Scholarship Council (file number: 201706010343). The funders had no role in study design, data collection and analysis, manuscript writing or decision to publish.

Disclosures: None.

References:

1. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research.* 2014;42:D1001-D1006.

2. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research.* 2017;45:D896-D901.

3. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet.* 2017;101:5-22.

4. Chen X, Wang Y. Tracking of blood pressure from childhood to adulthood: a systematic review and meta-regression analysis. *Circulation.* 2008;117:3171-3180.

5. Li Z, Snieder H, Harshfield GA, Treiber FA, Wang X. A 15-year longitudinal study on ambulatory blood pressure tracking from childhood to early adulthood. *Hypertens Res.* 2009;32:404-410.

6. Aarestrup J, Bjerregaard LG, Gamborg M, Angquist L, Tjonneland A, Overvad K, Linneberg A, Osler M, Mortensen EL, Gyntelberg F, *et al.* Tracking of body mass index from 7 to 69 years of age. *Int J Obes (Lond).* 2016;40:1376-1383.

7. Zimmermann E, Bjerregaard LG, Gamborg M, Vaag AA, Sorensen TIA, Baker JL. Childhood body mass index and development of type 2 diabetes throughout adult life-A large-scale danish cohort study. *Obesity (Silver Spring).* 2017;25:965-971.

8. Kupper N, Ge D, Treiber FA, Snieder H. Emergence of novel genetic effects on blood pressure and hemodynamics in adolescence: the Georgia Cardiovascular Twin Study. *Hypertension.* 2006;47:948-954.

9. Graff M, Ngwa JS, Workalemahu T, Homuth G, Schipf S, Teumer A, Volzke H, Wallaschofski H, Abecasis GR, Edward L, *et al.* Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. *Hum Mol Genet.* 2013;22:3597-3607.

10. Ntalla I, Panoutsopoulou K, Vlachou P, Southam L, William Rayner N, Zeggini E, Dedoussis GV. Replication of established common genetic variants for adult BMI and childhood obesity in Greek adolescents: the TEENAGE study. *Ann Hum Genet.* 2013;77:268-274.

11. Graae AS, Hollensted M, Kloppenborg JT, Mahendran Y, Schnurr TM, Appel EVR, Rask J, Nielsen TRH, Johansen MO, Linneberg A, *et al.* An adult-based insulin resistance genetic risk score associates with insulin resistance, metabolic traits and altered fat distribution in Danish children and adolescents who are overweight or obese. *Diabetologia.* 2018;61:1769-1779.

12. Oikonen M, Tikkanen E, Juhola J, Tuovinen T, Seppala I, Juonala M, Taittonen L, Mikkila V, Kahonen M, Ripatti S, *et al.* Genetic variants and blood pressure in a population-based cohort: the Cardiovascular Risk in Young Finns study. *Hypertension.* 2011;58:1079-1085.

13. Huisman M, Oldehinkel AJ, de Winter A, Minderaa RB, de Bildt A, Huizink AC, Verhulst FC, Ormel J. Cohort profile: the Dutch 'TRacking Adolescents' Individual Lives' Survey'; TRAILS. *Int J Epidemiol.* 2008;37:1227-1235.

14. Oldehinkel AJ, Rosmalen JG, Buitelaar JK, Hoek HW, Ormel J, Raven D, Reijneveld SA, Veenstra R, Verhulst FC, Vollebergh WA, *et al.* Cohort Profile Update: the TRacking Adolescents' Individual Lives Survey (TRAILS). *Int J Epidemiol.* 2015;44:76-76n.

15. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet.* 2018;27:3641-3649.

16. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, Yengo L, Ferreira T, Marouli E, Ji Y, *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet.* 2019;28:166-174.

17. Eppinga RN, Hagemeijer Y, Burgess S, Hinds DA, Stefansson K, Gudbjartsson DF, van Veldhuisen DJ, Munroe PB, Verweij N, van der Harst P. Identification of genomic loci

associated with resting heart rate and shared genetic predictors with all-cause mortality. *Nat Genet.* 2016;48:1557-1563.

18. van den Berg ME, Warren HR, Cabrera CP, Verweij N, Mifsud B, Haessler J, Bihlmeyer NA, Fu YP, Weiss S, Lin HJ, *et al.* Discovery of novel heart rate-associated loci using the Exome Chip. *Hum Mol Genet.* 2017;26:2346-2363.

19. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet.* 2018;50:1412-1425.

20. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, Tin A, Wang L, Chu AY, Hoppmann A, *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51:957-972.

21. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med.* 2017;14:e1002383.

22. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, Holm H, Sanna S, Kavousi M, Baumeister SE, *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet.* 2011;43:1131-1138.

23. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet.* 2010;42:105-116.

24. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Magi R, Strawbridge RJ, Rehnberg E, Gustafsson S, *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet.* 2012;44:991-1005.

25. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu CT, Bielak LF, Prokopenko I, *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet.* 2012;44:659-669.

26. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45:1274-1283.

27. Surakka I, Horikoshi M, Magi R, Sarin AP, Mahajan A, Lagou V, Marullo L, Ferreira T, Miraglio B, Timonen S, *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat Genet.* 2015;47:589-597.

28. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam D, Alves AC, *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet.* 2017;49:1758-1766.

29. Mack S, Coassin S, Rueedi R, Yousri NA, Seppala I, Gieger C, Schonherr S, Forer L, Erhart G, Marques-Vidal P, *et al.* A genome-wide association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a) isoforms. *J Lipid Res.* 2017;58:1834-1844.

30. Ligthart S, Vaez A, Vosa U, Stathopoulou MG, de Vries PS, Prins BP, Van der Most PJ, Tanaka T, Naderi E, Rose LM, *et al.* Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. *Am J Hum Genet.* 2018;103:691-706.

31. Granada M, Wilk JB, Tuzova M, Strachan DP, Weidinger S, Albrecht E, Gieger C, Heinrich J, Himes BE, Hunninghake GM, *et al.* A genome-wide association study of plasma total IgE concentrations in the Framingham Heart Study. *J Allergy Clin Immunol.* 2012;129:840-845.e821.

32. Felix JF, Bradfield JP, Monnereau C, van der Valk RJ, Stergiakouli E, Chesi A, Gaillard R, Feenstra B, Thiering E, Kreiner-Moller E, *et al.* Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet.* 2016;25:389-403.

33. Parmar PG, Taal HR, Timpson NJ, Thiering E, Lehtimaki T, Marinelli M, Lind PA, Howe LD, Verwoert G, Aalto V, *et al.* International Genome-Wide Association Study Consortium Identifies Novel Loci Associated With Blood Pressure in Children and Adolescents. *Circ Cardiovasc Genet.* 2016;9:266-278.

34. van Beek JH, Lubke GH, de Moor MH, Willemsen G, de Geus EJ, Hottenga JJ, Walters RK, Smit JH, Penninx BW, Boomsma DI. Heritability of liver enzyme levels estimated from genome-wide SNP data. *Eur J Hum Genet.* 2015;23:1223-1228.

35. van Beek JH, de Moor MH, de Geus EJ, Lubke GH, Vink JM, Willemsen G, Boomsma DI. The genetic architecture of liver enzyme levels: GGT, ALT and AST. *Behav Genet.* 2013;43:329-339.

36. Snieder H, van Doornen LJ, Boomsma DI. The age dependency of gene expression for plasma lipids, lipoproteins, and apolipoproteins. *Am J Hum Genet.* 1997;60:638-650.

37. Prins BP, Lagou V, Asselbergs FW, Snieder H, Fu J. Genetics of coronary artery disease: genome-wide association studies and beyond. *Atherosclerosis.* 2012;225:1-10.

38. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;460:748-752.

39. Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet.* 2015;97:576-592.

Table 1. Details on the transformations, covariates and exclusions used for genetic risk score analysis of the 20 selected traits in TRAILS

Trait	Abbreviation	transform	Covariates	Reference	TRAILS correction
Anthropometry					
Height	Height	INR	Sex, age	15	Yes
Body mass index	BMI	INR	Sex, age	15	Yes
Waist-to-hip ratio (BMI adjusted)	WHRadjBMI	INR	Sex, age, age^2, BMI	16	Yes
Cardiovascular and renal function					
Heart rate	HR		Sex, age, age^2, BMI	17,18	
Systolic blood pressure	SBP		Sex, age, age^2, BMI	19	Yes
Diastolic blood pressure	DBP		Sex, age, age^2, BMI	19	Yes
Estimated glomerular filtration rate	eGFR	ln	Sex, age	20	Yes
Metabolism					
Glycated hemoglobin*	HbA1c		Sex, age, age^2	21	Yes
Alanine transaminase	ALT	Log10	Sex, age	22	
Fasting glucose†	FG		Sex, age	23,24	Yes
Fasting glucose (BMI adjusted)†	FGadjBMI		Sex, age, BMI	24,25	Yes
Fasting insulin†	FI	ln	Sex, age	24	Yes
Fasting insulin (BMI adjusted)†	FIadjBMI	ln	Sex, age, BMI	24	Yes
High-density lipoprotein	HDL	INR	Sex, age, age^2	26-28	
Low-density lipoprotein	LDL	INR	Sex, age, age^2	26-28	
Total cholesterol	TC	INR	Sex, age, age^2	26-28	
Triglycerides	TG	ln, INR	Sex, age, age^2	26-28	
Lipoprotein(a)	Lp(a)	INR	Sex, age	29	
Inflammation					
C-reactive protein	CRP	ln	Sex, age	30	Yes
Immunoglobulin E	IgE	Log10	Sex, age	31	

*excluding individuals with diagnosed diabetes or high fasting glucose (≥7 mmol/l).
†excluding individuals with diagnosed diabetes or high fasting glucose (≥7 mmol/l) or non-fasting.
Abbreviations: ln, natural logarithm; INR, inverse normal of residuals.

Table 2. Descriptive statistics of age and the 20 quantitative traits at the third wave (16y) in the TRAILS cohort

Trait (unit)	total (n=1354)*	Male (n=644)*	Female (n=710)*
Age (year)	16.22 (0.66)	16.21 (0.64)	16.23 (0.68)
Anthropometry			
Height (cm)	174.58 (8.87)	180.27 (7.65)	169.29 (6.27)
Body mass index (kg/m^2)	20.75 (19.13-22.55)	20.27 (18.75-21.92)	21.25 (19.54-23.11)
Waist-to-hip ratio	0.83 (0.79-0.87)	0.83 (0.80-0.87)	0.83 (0.78-0.86)
Cardiovascular and renal function			
Heart rate (bpm)	68.03 (11.98)	66.40 (12.12)	69.54 (11.67)
Systolic blood pressure (mmHg)	118.29 (12.53)	122.33 (12.60)	114.57 (11.26)
Diastolic blood pressure (mmHg)	61.09 (6.95)	60.40 (7.06)	61.72 (6.79)
Estimated glomerular filtration rate† (mL/min per 1.73 m^2)	97.65 (89.28-107.41)	95.63 (86.91-105.73)	99.56 (91.41-108.32)
Metabolism			
HbA1c (%)	5.17 (0.45)	5.23 (0.47)	5.13 (0.42)
Alanine transaminase (U/I)	14.00 (12.00-18.00)	16.00 (13.00-20.00)	13.00 (11.00-16.00)
Fasting glucose (mmol/l)	4.54 (0.42)	4.61 (0.45)	4.47 (0.38)
Fasting insulin (mU/I)	12.00 (9.10-16.00)	11.05 (8.50-15.00)	12.00 (9.50-16.00)
High-density lipoprotein (mmol/l)	1.40 (1.20-1.60)	1.40 (1.20-1.60)	1.50 (1.30-1.70)
Low-density lipoprotein (mmol/l)	2.20 (1.80-2.60)	2.10 (1.70-2.50)	2.40 (2.00-2.79)
Total cholesterol (mmol/l)	3.70 (3.30-4.23)	3.50 (3.10-4.00)	4.00 (3.50-4.40)
Triglycerides (mmol/l)	0.69 (0.52-0.93)	0.64 (0.49-0.90)	0.72 (0.55-0.96)
Lipoprotein(a) (mg/l)	69.50 (30.25-220.50)	60.00 (25.00-175.00)	81.00 (35.00-260.00)
Inflammation			
C-reactive protein (mg/l)	0.40 (0.20-1.00)	0.30 (0.20-0.80)	0.50 (0.20-1.40)
Immunoglobulin E (kU/I)	67.05 (22.23-213.00)	67.20 (22.70-225.00)	66.30 (21.90-205.00)

*Descriptives are either mean (SD) or median (interquartile range) depending on the distribution of the variable.
†eGFR=41.3 * (height/SCr), SCr, serum creatinine in mg/dl.
T1 – T5, wave 1 – 5. The number of participants of the five waves were 1354, 1349, 1346, 1291, 1259, respectively.

Table 3. The result of genetic risk scores analyses at the third wave (16 y)

Trait	N in TRAILS	Number of SNPs	variance explained (wGRS) in TRAILS adolescents		variance explained in adults (%)*	difference of traits between top and bottom wGRS decile†
			R^2 (%)	P		
Anthropometry						
Height	1292	3290	13.68	1.41E-71	19.70	10.82 (9.32, 12.32)***
BMI	1289	941	6.47	3.59E-21	5.00	2.21 (1.46, 2.97)***
WHRadjBMI	1285	462	1.38	9.26E-06	3.90	0.03 (0.02, 0.05)***
Cardiovascular and renal function						
HR	1280	80	1.46	9.39E-06	2.50	6.50 (3.66, 9.35)***
SBP	1280	970	2.15	5.22E-09	5.70	6.30 (3.54, 9.07)***
DBP	1280	962	4.48	1.14E-14	5.32	5.61 (3.93, 7.28)***
eGFR	1074	253	5.04	4.28E-14	7.01	10.85 (7.40, 14.31)***
Metabolism						
HbA1C	1074	43	2.83	1.62E-08	4.19	0.29 (0.18, 0.40)***
ALT	1082	4	0.10	2.86E-01	0.10	0.01 (-0.03, 0.06)
FG	978	31	3.67	1.01E-09	3.28	0.25 (0.15, 0.36)***
FGadjBMI	968	19	0.95	2.06E-03		0.25 (0.15, 0.36)***
FI	969	12	1.45	1.37E-04	1.20	0.28 (0.15, 0.40)***
FIadjBMI	959	12	0.69	5.70E-03		0.23 (0.11, 0.34)***
HDL	1082	247	11.49	1.70E-32	12.80	0.36 (0.28, 0.43)***
LDL	1082	194	18.49	2.38E-52	19.50	1.04 (0.89, 1.20)***
TC	1082	234	12.95	3.50E-38	18.80	0.94 (0.77, 1.12)***
TG	1082	190	6.56	4.63E-18	9.30	0.47 (0.36, 0.58)***
LPA	1079	49	39.59	4.94E-123	36.00	39.76 (35.78, 43.75)***
Inflammation						
CRP	1078	77	3.69	8.90E-11	11.00	0.82 (0.51, 1.14)***
IgE	1060	7	2.06	2.52E-06		0.36 (0.18, 0.54)***

see Table 1 for abbreviations.
* These results were extracted from the literature.
† The transformations and unit of phenotypes: height (cm), BMI (Kg/m^2), HR (bpm), SBP (mmHg), DBP (mmHg), eGFR (mL/min per 1.73 m^2), HbA1C (%), ALT (U/I, log10 transformation), FG (mmol/l), FGadjBMI (mmol/l), FI (mmol/l, ln transformation), FIadjBMI (mmol/l, ln transformation), HDL (mmol/l), LDL (mmol/l), TC (mmol/l), TG (mmol/l, ln transformation), LPA (mg/l), CRP (mg/l, ln transformation), IgE (kU/l, log10 transformation).
****P*<0.001.

Figure Legends:

Figure 1. Flowchart showing the process and results of SNP selection of the 20 traits of interest

Figure 2. Variance explained by uGRSs and wGRSs for anthropometric traits at different ages.

Figure 3. The comparison between variances explained in adolescents and in adults. At the red dashed line the variances explained in adolescents and adults are the same. Seventeen traits are shown: height, BMI, WHR, HR, SBP, DBP, eGFR, HbA1c, ALT, FG, FI, HDL, LDL, TC, TG, Lp(a), CRP (see Table 1 for abbreviations). For some traits our results were not completely comparable with those from literature as different methods compared to ours were used for some traits to estimate variance explained in adults. In the literature, the method that included all SNPs into a linear regression model, adjusted for covariates and calculated the adjusted R^2 was used for WHRadjBMI, SBP, DBP, HbA1c, FI, HDL, LDL, TC, and TG. The formula ((2*MAF(1-MAF)b^2)/var) was used for eGFR and CRP. In addition, for some traits not exactly the same SNPs as ours were included to evaluate variance explained in adults. Some traits included a few more SNPs than ours (SBP, DBP, eGFR, HbA1c, FI), while some traits included a few less (WHRadjBMI, HR, HDL, LDL, TC, TG, Lp(a)). See Supplementary Table 26 for more details.

Figure 4. Odds ratios of overweight/obesity and hypertension comparing each of the upper nine GRS deciles with the lowest decile. Deciles of wGRS for BMI was used for overweight/obesity, and deciles of wGRS for SBP was used for hypertension (as most cases of hypertension resulted from high SBP).